

## Melakukan Uji Validitas, Reliabilitas, Daya Beda, dan Taraf Kesukaran Soal

Inayatul Ulya<sup>1</sup> Reny Atika<sup>2</sup>

Pendidikan Agama Islam, Tarbiyah dan Keguruan, Institut Agama Islam Negeri Datuk  
Laksemana Bengkulu, Indonesia<sup>1,2</sup>

Email: [ulyaaasr@gmail.com](mailto:ulyaaasr@gmail.com)<sup>1</sup> [renyatyka@gmail.com](mailto:renyatyka@gmail.com)<sup>2</sup>

### Abstrak

Evaluasi pembelajaran merupakan bagian penting dalam proses pendidikan yang bertujuan untuk mengetahui tingkat keberhasilan peserta didik dalam mencapai tujuan pembelajaran. Kualitas hasil evaluasi sangat dipengaruhi oleh kualitas instrumen yang digunakan, sehingga diperlukan analisis butir soal melalui uji validitas, reliabilitas, daya beda, dan taraf kesukaran. Penelitian ini menggunakan metode studi pustaka (library research) dengan pendekatan deskriptif kualitatif. Data diperoleh dari berbagai sumber literatur seperti buku, jurnal, dan artikel ilmiah yang berkaitan dengan analisis butir soal. Teknik pengumpulan data dilakukan melalui dokumentasi dan kajian literatur, kemudian dianalisis secara deskriptif untuk menjelaskan konsep, tujuan, serta prosedur analisis validitas, reliabilitas, daya beda, dan taraf kesukaran soal. Hasil kajian menunjukkan bahwa instrumen evaluasi yang baik harus memenuhi kriteria valid, reliabel, mampu membedakan tingkat kemampuan peserta didik, serta memiliki tingkat kesukaran yang proporsional. Oleh karena itu, analisis butir soal sangat penting untuk meningkatkan kualitas instrumen evaluasi agar hasil penilaian lebih akurat, objektif, dan dapat dipertanggungjawabkan.

**Kata Kunci:** Analisis Butir Soal, Validitas, Reliabilitas, Daya Beda, Taraf Kesukaran



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

## PENDAHULUAN

Evaluasi pembelajaran merupakan bagian yang tidak terpisahkan dari keseluruhan proses pendidikan. Kegiatan ini memiliki peran strategis dalam mengukur keberhasilan peserta didik dalam mencapai tujuan pembelajaran yang telah ditetapkan. Melalui evaluasi, pendidik tidak hanya memperoleh informasi mengenai hasil belajar, tetapi juga dapat menilai efektivitas metode, strategi, serta perangkat pembelajaran yang digunakan. Oleh karena itu, evaluasi pembelajaran harus dilaksanakan secara sistematis, objektif, dan berkelanjutan agar mampu memberikan gambaran yang akurat mengenai capaian belajar peserta didik. Salah satu instrumen yang paling sering digunakan dalam evaluasi pembelajaran adalah tes. Tes digunakan untuk mengukur kemampuan kognitif peserta didik melalui berbagai bentuk soal, seperti pilihan ganda, benar-salah, maupun uraian. Namun demikian, kualitas suatu tes sangat ditentukan oleh kualitas butir soal yang menyusunnya. Soal yang disusun tanpa melalui analisis yang baik berpotensi menghasilkan data yang bias, tidak akurat, bahkan menyesatkan dalam pengambilan keputusan pendidikan. Hal ini tentu akan berdampak pada ketidaktepatan dalam menilai kemampuan peserta didik serta kurang optimalnya perbaikan proses pembelajaran.

Dalam kenyataannya, masih banyak pendidik yang menyusun soal hanya berdasarkan intuisi atau pengalaman, tanpa melakukan analisis empiris terhadap kualitas butir soal tersebut. Akibatnya, soal yang digunakan sering kali belum memenuhi kriteria sebagai instrumen evaluasi yang baik. Padahal, instrumen evaluasi yang berkualitas harus memenuhi beberapa persyaratan utama, di antaranya validitas, reliabilitas, daya beda, dan taraf kesukaran. Keempat aspek ini merupakan indikator penting yang menentukan apakah suatu soal layak digunakan atau perlu direvisi. Uji validitas dilakukan untuk mengetahui sejauh mana suatu butir soal mampu mengukur apa yang seharusnya diukur. Dengan kata lain, validitas

berkaitan dengan ketepatan fungsi suatu instrumen. Sementara itu, uji reliabilitas berkaitan dengan tingkat konsistensi hasil pengukuran, yaitu apakah suatu instrumen dapat memberikan hasil yang relatif sama jika digunakan dalam kondisi yang serupa. Instrumen yang reliabel akan menghasilkan data yang stabil dan dapat dipercaya.

Selain itu, daya beda soal menjadi aspek penting dalam analisis butir soal karena berkaitan dengan kemampuan soal dalam membedakan peserta didik yang memiliki kemampuan tinggi dan rendah. Soal yang baik seharusnya dapat dijawab dengan benar oleh peserta didik yang berkemampuan tinggi dan cenderung salah dijawab oleh peserta didik yang berkemampuan rendah. Di sisi lain, taraf kesukaran soal digunakan untuk mengetahui tingkat kesulitan suatu butir soal, apakah termasuk kategori mudah, sedang, atau sulit. Soal yang terlalu mudah atau terlalu sulit tidak akan mampu memberikan informasi yang optimal mengenai kemampuan peserta didik. Keempat aspek tersebut saling berkaitan dan harus dianalisis secara menyeluruh untuk menghasilkan instrumen evaluasi yang berkualitas. Dengan melakukan analisis validitas, reliabilitas, daya beda, dan taraf kesukaran, pendidik dapat mengidentifikasi kelemahan pada butir soal serta melakukan perbaikan yang diperlukan. Hal ini penting agar instrumen yang digunakan benar-benar mampu mengukur kemampuan peserta didik secara tepat, adil, dan objektif. Berdasarkan latar belakang tersebut, artikel ini disusun untuk membahas secara komprehensif mengenai uji validitas, reliabilitas, daya beda, dan taraf kesukaran soal. Selain itu, artikel ini juga memaparkan konsep dasar, tujuan, serta cara perhitungan dari masing-masing uji tersebut. Dengan adanya pembahasan ini, diharapkan dapat memberikan pemahaman yang lebih mendalam kepada pendidik maupun peneliti dalam menyusun dan menganalisis instrumen evaluasi pembelajaran, sehingga kualitas penilaian yang dihasilkan menjadi lebih baik dan dapat dipertanggungjawabkan secara ilmiah.

## METODE PENELITIAN

Penelitian ini menggunakan metode studi pustaka (*library research*) dengan pendekatan deskriptif kualitatif. Data penelitian diperoleh dari berbagai sumber literatur seperti buku, jurnal ilmiah, artikel, dan referensi lain yang berkaitan dengan uji validitas, reliabilitas, daya beda, dan taraf kesukaran soal. Teknik pengumpulan data dilakukan melalui dokumentasi dan kajian literatur, kemudian dianalisis secara deskriptif dengan cara mengidentifikasi, mengklasifikasikan, serta menginterpretasikan berbagai konsep dan teori yang relevan. Melalui metode ini, penelitian bertujuan untuk memberikan pemahaman mengenai pentingnya analisis butir soal dalam meningkatkan kualitas instrumen evaluasi pembelajaran agar lebih valid, reliabel, objektif, dan dapat dipertanggungjawabkan.

## HASIL PENELITIAN DAN PEMBAHASAN

### Pengertian Uji Validitas, Reliabilitas, Daya Beda dan Taraf Kesukaran Soal

#### *Uji Validitas*

Uji validitas merupakan uji yang berfungsi untuk melihat apakah suatu alat ukur tersebut valid (sahih) atau tidak valid. Alat ukur yang dimaksud disini merupakan pertanyaan-pertanyaan yang ada dalam kuesioner. Dalam uji pengukuran validitas terdapat dua macam yaitu pertama, mengkorelasikan antar skor butir pertanyaan (item) dengan total item. Kedua, mengkorelasikan antar masing-masing skor indikator item dengan total skor konstruk. Uji validitas juga merupakan derajat yang menunjukkan suatu tes mengukur apa yang hendak diukur. Sebuah data untuk mencari *r hitung* dari setiap item pertanyaan kuesioner. Dalam perspektif Islam, konsep validitas juga berkaitan dengan kebenaran dan ketepatan dalam menilai. Sebagaimana firman Allah dalam Al-Qur'an: "*Wahai orang-orang yang beriman! Jika seseorang yang fasik datang kepadamu membawa suatu berita, maka telitilah kebenarannya...*"

(QS. Al-Hujurat: 6). Menurut Ibnu Katsir, QS. Al-Hujurat ayat 6 menjelaskan bahwa Allah memerintahkan kaum mukmin untuk berhati-hati dalam menerima berita dari orang fasik dengan cara meneliti dan memeriksa terlebih dahulu kebenarannya, agar tidak terjadi kesalahan dalam mengambil keputusan yang dapat menimbulkan penyesalan di kemudian hari. Selain itu menurut M. Quraish Shihab, QS. Al-Hujurat ayat 6 mengandung perintah untuk melakukan tabayyun, yaitu meneliti dan memastikan kebenaran suatu informasi sebelum mempercayai atau menyebarkannya, agar tidak terjadi kesalahan, fitnah, maupun kerugian kepada pihak lain.

### **Reliabilitas**

Uji Reliabilitas merujuk pada satu pengertian bahwa suatu instrumen cukup dapat dipercaya untuk digunakan sebagai alat pengumpul data karena instrumen tersebut sudah baik. Uji reliabilitas menurut Husaini, adalah proses pengukuran terhadap ketepatan (konsisten) dari suatu instrumen. Uji reliabilitas dapat digunakan untuk mengetahui sebuah konsistensi alat ukur, apakah dapat alat yang digunakan dengan konsisten saat diuji berulang. Alat ukur yang dapat dikatakan reliabel apabila Ketika dilakukan uji berulang akan mendapat hasil yang konsisten dan sama. Sebelum dilakukan uji reliabilitas, dilakukan uji validitas yang bertujuan untuk mengetahui apakah data tersebut valid sebelum dilakukannya reliabilitas. Jika data yang telah melalui proses validitas tidak valid, maka perlu dilakukannya uji reliabilitas sebab reliabilitas dapat atau bisa dilakukan Ketika data yang melalui proses validasi.

### **Daya Beda**

Daya beda sering juga disebut dengan daya diskriminasi. Dalam hal ini, pembuat tes atau *tester* (orang yang melaksanakan tes atau orang yang ditugasi untuk melaksanakan pengambilan tes terhadap responden, misalkan guru atau dosen) bermaksud untuk mendiskriminasikan atau membedakan hasil belajar peserta tes (pihak yang dikenai tes, misalkan siswa atau mahasiswa) yang berkemampuan tinggi dengan peserta yang berkemampuan rendah.

### **Taraf Kesukaran Soal**

Uji taraf kesukaran dilakukan untuk melihat sukar atau tidaknya suatu soal. Arikunto menyatakan bahwa "soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar." Berdasarkan dilakukan uji taraf kesukaran soal uji coba, terdapat 12 soal dengan kriteria mudah, 24 soal dengan kriteria sedang, dan 4 soal dengan kriteria sulit. Uji Tingkat kesukaran pada soal bertujuan untuk mengukur sejauh mana Tingkat kesulitan pada setiap butir soal yang telah diberikan kepada siswa.<sup>1</sup> Berdasarkan pengertian di atas, penulis menyimpulkan bahwa melakukan uji validitas, reliabilitas, daya beda, dan taraf kesukaran soal adalah suatu proses analisis terhadap butir soal yang bertujuan untuk menilai kualitas instrumen evaluasi pembelajaran secara menyeluruh, sehingga soal yang digunakan benar-benar mampu mengukur kemampuan peserta didik secara tepat, konsisten, mampu membedakan tingkat kemampuan, serta memiliki tingkat kesulitan yang sesuai. Dengan demikian, instrumen evaluasi yang dihasilkan menjadi lebih akurat, objektif, dan dapat dipertanggungjawabkan.

<sup>1</sup> Najwa Zalfa Zuhri and Syihabuddin, *Analisis Validitas, Reliabilitas, Dan Tingkat Kesukaran Soal Bahasa Arab Tingkat SMP Berbasis Artificial Inlegence(AI) Melalui Platform Question Well*, (Bandung), Jurnal Pendidikan dan Pembelajaran Indonesia (JPPI), vol.4 no. 2 (2024): hal.699.

---

## **Tujuan Uji Validitas, Reliabilitas, Daya Beda dan Taraf Kesukaran Soal**

### ***Uji validitas***

Uji validitas digunakan untuk menentukan seberapa akurat pernyataan yang akan dibuat kepada responden dalam kuesioner. Validitas suatu alat ukur dapat ditentukan dengan suatu tes yang disebut dengan uji validitas. Analisis uji validitas bertujuan untuk mengetahui sejauh mana suatu butir soal mampu mengukur apa yang seharusnya diukur dalam suatu tes. Melalui uji validitas, pendidik dapat menilai ketepatan dan kesesuaian setiap butir soal dengan indikator pembelajaran yang telah ditetapkan. Selain itu, uji validitas juga berfungsi untuk mengidentifikasi butir soal yang layak digunakan maupun yang perlu diperbaiki atau dibuang karena tidak memenuhi kriteria. Dengan demikian, instrumen evaluasi yang digunakan benar-benar mampu menghasilkan data yang akurat dan mencerminkan kemampuan peserta didik secara tepat, sehingga hasil penilaian dapat dipertanggungjawabkan secara ilmiah.

### ***Reliabilitas***

Uji reliabilitas digunakan untuk mengetahui sebuah konsistensi dari alat ukur, apakah dapat alat yang digunakan dengan konsisten saat diuji berulang. Alat ukur yang dapat dikatakan reliabel apabila ketika dilakukan uji berulang akan mendapat hasil yang konsisten dan sama. Sebuah kuesioner dapat dikatakan reliabel apabila dalam proses pengambilan datanya dilakukan berulang-ulang dapat menghasilkan nilai dan pengaruh yang sama dan sebaliknya apabila kuesioner tidak reliabel maka dalam pengambilan data secara berulang-ulang akan tidak sama dengan hasil yang didapat terakhir kali saat pengujian data. Jadi, dapat diartikan bahwa uji reliabilitas ialah sebuah alat ukur yang memiliki konsistensi dalam pengambilan, pengumpulan, dan hasil akhir yang reliabel dengan pembuktian bahwa alat, bahan, serta data yang diuji memiliki konsistensi yang sama dalam proses dan tahap pengambilan data yang dilakukan secara berulang-ulang dan tidak dapat berubah.

### ***Daya Beda***

Daya beda suatu butir tes bertujuan untuk menentukan dapat tidaknya suatu butir tes berfungsi untuk menentukan dapat tidaknya suatu butir tes membedakan kelompok dalam aspek yang diukur sesuai dengan perbedaan yang ada pada kelompok itu. Tujuan dari penelaahan daya beda adalah untuk melihat kemampuan butir tes tertentu dalam membedakan antara pengambilan tes yang berkemampuan tinggi dan pengambil tes yang berkemampuan rendah.

### ***Taraf Kesukaran Soal***

Analisis taraf kesukaran soal bertujuan untuk mengetahui tingkat kesulitan setiap butir soal dalam suatu tes, sehingga dapat diklasifikasikan ke dalam kategori mudah, sedang, atau sulit. Melalui analisis ini, pendidik dapat menilai kualitas soal yang digunakan dalam evaluasi pembelajaran serta memastikan kesesuaian antara tingkat kesulitan soal dengan kemampuan peserta didik. Selain itu, analisis taraf kesukaran juga berfungsi sebagai dasar dalam memperbaiki atau merevisi butir soal yang kurang sesuai, serta membantu dalam menyusun komposisi soal yang seimbang agar mampu mengukur berbagai tingkat kemampuan peserta didik. Dengan demikian, hasil evaluasi yang diperoleh menjadi lebih akurat, objektif, dan dapat digunakan sebagai dasar pengambilan keputusan dalam proses pembelajaran.

## Analisis Perhitungan Uji Validitas, Reliabilitas, Daya Beda, dan Taraf Kesukaran Soal

### Uji Validitas Soal

Uji validitas digunakan untuk mengetahui apakah suatu butir soal benar-benar mengukur apa yang seharusnya diukur dalam sebuah tes.

#### Korelasi Pearson Product Moment

$$r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

#### Keterangan:

$r_{xy}$  = koefisien validitas butir soal

$X$  = skor tiap butir soal

$Y$  = skor total

$N$  = jumlah peserta tes

$\sum$  = jumlah keseluruhan data

Rumus ini digunakan untuk melihat hubungan antara skor butir soal dengan skor total tes. Semakin tinggi korelasi, semakin valid soal tersebut.

#### Korelasi Point Biserial

$$r_{pb} = \frac{M_p - M_t}{S_t} \sqrt{\frac{p}{q}}$$

#### Keterangan:

$r_{pb}$  = koefisien validitas point biserial

$M_p$  = rata-rata skor total siswa yang menjawab benar

$M_t$  = rata-rata skor total seluruh siswa

$S_t$  = standar deviasi skor total

$p$  = proporsi jawaban benar

$q$  = proporsi jawaban salah ( $q = 1 - p$ )

Rumus ini digunakan untuk soal berbentuk benar-salah untuk melihat perbedaan kemampuan kelompok yang menjawab benar dan salah. Dari berbagai rumus validitas, rumus Pearson Product Moment paling sering digunakan karena fleksibel untuk berbagai jenis data dan mudah dihitung baik manual maupun dengan software statistic

### Uji Reliabilitas Soal

Reliabilitas adalah tingkat konsistensi suatu instrumen dalam menghasilkan skor yang stabil pada pengukuran yang berulang.

#### KR-20

$$r_{11} = \frac{k}{k-1} \left( 1 - \frac{\sum pq}{S^2} \right)$$

**Keterangan:** $r_{11}$  = reliabilitas tes $k$  = jumlah butir soal $p$  = proporsi siswa menjawab benar $q$  = proporsi siswa menjawab salah ( $q = 1 - p$ ) $S^2$  = varians total skor $pq$  = variasi tiap butir soal

KR-20 digunakan untuk tes pilihan ganda benar-salah untuk melihat konsistensi antar butir soal.

*KR-21*

$$r_{11} = \frac{k}{k-1} \left( 1 - \frac{M(k-M)}{kS^2} \right)$$

**Keterangan:** $M$  = rata-rata skor total $k$  = jumlah butir soal $S^2$  = varians total skor

KR-21 digunakan jika data tidak lengkap, tetapi hasilnya kurang akurat dibanding KR-20.

*Alpha Cronbach*

$$\alpha = \frac{k}{k-1} \left( 1 - \frac{\sum S_i^2}{S_t^2} \right)$$

**Keterangan:** $\alpha$  = koefisien reliabilitas $k$  = jumlah item $S_i^2$  = varians tiap item $S_t^2$  = varians total

Alpha Cronbach digunakan untuk instrumen non-dikotomus seperti angket, essay, atau skala sikap. Dari beberapa rumus reliabilitas, yang paling banyak di gunakan adalah Alpha Cronbach, hal ini dikarenakan dapat dipakai untuk berbagai jenis instrumen penelitian dan hasilnya lebih stabil.

**Daya Beda Soal**

Daya beda menunjukkan kemampuan soal untuk membedakan siswa berkemampuan tinggi dan rendah.

*Indeks Daya Beda (D)*

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B}$$

**Keterangan:** $D$  = indeks daya beda $B_A$  = jumlah siswa kelompok atas yang benar

$B_B$  = jumlah siswa kelompok bawah yang benar

$J_A$  = jumlah siswa kelompok atas

$J_B$  = jumlah siswa kelompok bawah

Rumus ini membandingkan kemampuan kelompok atas dan bawah dalam menjawab soal dengan benar.

### **Taraf Kesukaran Soal**

Taraf kesukaran adalah tingkat mudah atau sulitnya suatu soal berdasarkan proporsi siswa yang menjawab benar.

*Indeks Kesukaran (P)*

$$P = \frac{B}{JS}$$

### **Keterangan:**

$P$  = indeks kesukaran

$B$  = jumlah siswa yang menjawab benar

$JS$  = jumlah seluruh siswa

Rumus ini menunjukkan tingkat kesulitan soal berdasarkan banyaknya siswa yang menjawab benar.

### **KESIMPULAN**

Berdasarkan hasil pembahasan, dapat disimpulkan bahwa kualitas suatu instrumen tes sangat ditentukan oleh empat aspek utama, yaitu validitas, reliabilitas, daya beda, dan taraf kesukaran. Uji validitas digunakan untuk memastikan bahwa soal mampu mengukur apa yang seharusnya diukur, sedangkan reliabilitas menunjukkan tingkat konsistensi hasil pengukuran. Daya beda berfungsi untuk membedakan peserta didik yang memiliki kemampuan tinggi dan rendah, sementara taraf kesukaran menunjukkan tingkat kesulitan suatu soal. Soal yang baik adalah soal yang memiliki validitas tinggi, reliabilitas yang baik, daya beda yang efektif, serta tingkat kesukaran yang seimbang (tidak terlalu mudah atau terlalu sulit). Dengan melakukan analisis terhadap keempat aspek tersebut, pendidik dapat memperbaiki kualitas butir soal sehingga instrumen evaluasi yang digunakan menjadi lebih akurat, objektif, dan dapat dipertanggungjawabkan dalam mengukur hasil belajar peserta didik.

### **DAFTAR PUSTAKA**

- Arikunto, S. (2019). Dasar-Dasar Evaluasi Pendidikan. Bumi Aksara.
- Forester, J., Brayen., Idris. A. (2024). Penelitian Kuantitatif: Uji Reliabilitas. (Palembang), Jurnal Pendidikan, Ilmu Sosial, dan Pengabdian Kepada Masyarakat, Vol. 4, No, 3.
- Istiyono. (2024). Prinsip-Prinsip Pengukur & Evaluasi Pendidikan. CV. Ruang Tentor.
- Katsir, I. (2008). Tafsir Ibnu Katsir, Jilid 7, Terj. M. Abdul Ghoffar. Pustaka Imam Asy-Syafi'i. Kementerian Agama Republik Indonesia. Al-Qur'an Dan Terjemahannya.
- Limbong, S. P., Syah. (2022). Model Buku Pembelajaran Ctl Dan Rme Terhadap Hasil Belajar Matematika Siswa. Guepedia.
- Ramdani., Dadi., Supriatna, E. (2023). Validitas Dan Reliabilitas Angket Kematangan Emosi. 6 No. 3.
- Saputri, A., Hera, And Zulhijrah. (2023). Analisis Instrumen Assesmen: Validitas, Reliabilitas, Tingkat Kesukaran Dan Daya Beda Butir Soal. (Yogyakarta), Jurnal Ilmiah PGSD FKIP Universitas Mandiri, Vol. 09 No. 05.



- Setyaedhi, S. Hari, Rusijono. (2025). Uji T: Uji Komparatif Dua Parameter Rata-Rata (Perhitungan Manual Dan Spss). Nas Media Pustaka.
- Shihab, M. K. (2002). Tafsir Al-Mishbah: Pesan, Kesan, Dan Keserasian Al-Qur'an, Vol. 12. Lentera Hati.
- Sudijono, A. (2020). Pengantar Evaluasi Pendidikan. RajaGrafindo Persada, 2020.
- Sugiyono. (2015). Metode Penelitian Pendidikan. Alfabeta.
- Zuhri., Najwa., Syihabuddin. (2024). Analisis Validitas, Reliabilitas, Dan Tingkat Kesukaran Soal Bahasa Arab Tingkat SMP Berbasis Artificial Inlegence(AI) Melalui Platform Question Well. (Bandung), Jurnal Pendidikan dan Pembelajaran Indonesia (JPPI), vol.4 no. 2.